Christian Linz
linz@cg.cs.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Prof. Dr. Ing. Marcus Magnor
magnor@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Dense Correspondence Estimation for Image Interpolation

Technical Report 2010-11-13

November 26, 2010

Computer Graphics Lab, TU Braunschweig

1 Abstract

We evaluate the current state-of-the-art in dense correspondence estimation for the use in multi-image interpolation algorithms. The evaluation is carried out on three real-world scenes and one synthetic scene, each featuring varying challenges for dense correspondence estimation. The primary focus of our study is on the perceptual quality of the interpolation sequences created from the estimated flow fields. Perceptual plausibility is assessed by means of a psychophysical userstudy. Our results show that current state-of-the-art in dense correspondence estimation does not produce visually plausible interpolations.

2 Introduction

Dense correspondence fields lie at the very heart of multi-image interpolation techniques. In literature, there are numerous approaches to compute such dense correspondence fields from a given image pair. Many of the more recent approaches are ranked in the Middlebury flow evaluation database [7], a database nowadays used to evaluate the performance of correspondence estimation techniques on a set of test scenes. While this database also evaluates interpolation quality by measuring an interpolation error, its primary focus is on measuring the accurracy of the obtained correspondence fields. However, the evaluation of a single frame halfway between two images does not capture the aspect of temporal coherence which is of great importance to image interpolation techniques.

In Refs. [53] and [52], Stich et al. compared their approach to interpolation sequences generated from dense correspondence fields computed with well-known optical flow algorithms [27, 15] by means of a psychophysical userstudy. While the proposed edge-based correspondence estimation technique yields perceptually higher quality interpolation sequences, Stich et al. do not compare their results against interpolation sequences generated from motion fields of state-of-the-art optical flow research as reflected by the Middlebury evaluation database [7].

In this report, we evaluate the current state-of-the-art in optical flow research for image interpolation sequences. In contrast to the Middlebury evaluation database [7], our focus is not on numerical accuracy of the obtained flow fields, but on perceptual plausibility of the interpolation sequences. The rest of this report is structured as follows: we start with an in-depth review of the state-of-the-art, Sect. 3. This is followed by a brief overview on applications of optical flow algorithms in a computer graphics context in Sect. 4. After that, we evaluate the current state-of-the-art in optical flow algorithms on several real-world scenes used in the Virtual Video Camera system [33] in Sect. 5. We summarize our evaluation and draw conclusions

in Sect. 6.

3 Related Work

Dense correspondence estimation has a long standing history in computer vision research, and a huge number of papers on different aspects of the problem have been published. Since a complete survey on all optical flow algorithms is out of the scope of this report, we refer the interested reader to Refs. [1, 8, 41, 10, 39, 54, 38] for previous surveys on the state-of-the-art instead. In the following sections, we will focus on more recent optical flow methods, especially on those ranked on the Middlebury optical flow page [7] since those can safely be considered to represent current state-of-the-art. Nevertheless, we also include the discussion of two older methods, i.e. Horn-Schunck and Lucas-Kanade, for their popularity in the computer graphics community and their availability in several open source libraries. We follow a classification introduced in Ref. [3] and extend it suitably.

3.1 Differential methods

Differential methods are based on a first degree approximation of the brightness constancy assumption, i.e.

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} = 0. \tag{1}$$

Horn-Schunck. Eq. (1) can be solved globally by adding an additional regularization term to the under-determined system and enforcing global smoothness. This is the classic Horn-Schunck approach [27]. Their idea is to minimize

$$\int_\Omega \left((\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t})^2 + \alpha(|\nabla \delta x|^2 + |\nabla \delta y|^2) \right) dxdy.$$

Ω denotes the image domain and α weighs the influence of the smoothness term.

Lucas-Kanade. Lucas and Kanade [35] take a different approach: they solve Eq. (1) for every pixel by assuming all pixels (x, y) within a fixed-size window move with the same flow and construct an over-determined system of equations,

$$w(x,y)\left(\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\right)\Big|_{x,y,t} = 0.$$

Since the system is over-determined, δx and δy are found as a weighted least squares solution. The weights $w(\cdot)$ diminish the influence of pixels farther away.

Combined local-global approach. Often, local methods such as Lucas-Kanade are more robust against noise, while global methods such as Horn-Schunck yield dense flow fields inside homogeneous regions. Bruhn et al. proposed a combined local-global (CLG) approach that incorporates the advantages of both paradigms [16], : It is highly robust under Gaussian noise while giving dense flow fields.

3.2 Multi-scale

The linear approximation made in Eq. (1) is only valid for small displacements. A common solution to estimate larger displacements is to use a multi-resolution coarse-to-fine approach. An image pyramid is constructed by repeatedly downsampling the image [23, 9]. The optical flow can then be computed on the coarsest resolution, and an upsampled version of δx and δy is used to initialize the solution on the next finer level. This process is then iterated until the final image resolution is reached.

Usually, the image is downsampled by a factor of two. It has long been believed that the performance of multi-resolution algorithms can be further improved by controlling the spatial frequency content more finely than with power of two image pyramids. This is usually achieved by filtering the individual pyramid level with a low-pass filter to generate different scales [15]. However, Sun et al. recently showed that the influence of the downsampling factor on flow quality does not have any statistical significance [55].

Further, multi-scale methods are limited by object size: if the object motion is larger than the extent of the object itself, the object will be smoothed away before its motion can be estimated.

3.3 Variational methods

Recent work on optical flow has mostly turned to variational approaches, seeking to minimize an energy functional over the entire image domain Ω. While differential methods linearize the optical flow equation already in the problem formulation, Brox et al. perform a non-linear optimization by postponing linearization to the numerical scheme [15]. Expressed in equation form, the functional reads

$$E(\mathbf{v}) = \int_\Omega (\underbrace{\sum_i \beta_i \Psi_i(D_i)}_{\text{data term}} + \alpha \underbrace{\Phi(\mathbf{v}, \nabla\mathbf{v}, \ldots)}_{\text{smoothness term}}) dxdy,$$

where D_i denotes the data term, usually some function of the input images, Ψ_i and Θ are robust functions to penalize outliers.

Data terms. The data term typically takes the form

$$\Psi_i(D_i) = \Psi_i(\mathcal{L}_i(I(x+\delta x, y+\delta y, t+\delta t)) - \mathcal{L}_i(I(x,y,t))).$$

The linear function $\mathcal{L}_i$ (e.g., identity, gradient, Laplacian) allows to generalize the gradient constancy assumption to include other constancy assumptions as well. Substituting $\Psi(x) = x^2$ and assuming the identity function for $\mathcal{L}_i$, one arrives at the classical Horn-Schunck data term. This data term has been extended by Brox et al. to also include the image gradient in the data term [15]. Going beyond intensity and gradient differences, Wedel et al. also integrated the fundamental matrix F into the data term [63]. If the scene is static, this helps to further restrict possible flow vectors in texture-less regions. While the classical approaches use the L^2-norm as penalty function in the data term, Brox et al. propose to use the robust Charbonnier penalty $\Psi(s^2) = \sqrt{(s^2 + \epsilon^2)}$ to robustify flow computation [15]. Recently, researches also tend to use $\Psi(x) = |x|$ in the data term [69, 66]. This robustifies the flow computation against illumination changes, occlusion and noise. Unfortunately, this norm is not differentiable and thus difficult to use.

Methods of regularization. Much research has been devoted to finding a good regularization strategy to fill in missing regions. The classical approach of Horn-Schunck is to use the L^2-norm of the flow field gradient as regularizer. It yields a convex functional that can be optimized globally and strongly penalizes discontinuities in the optical flow field. This usually leads to blurry flow fields around edges. Starting with the approach of Black and Anandan [11], different robust functions have been proposed for regularization. However, some of them are non-convex and thus difficult to optimize.

Based on this approach, Zach et al. propose to use a true total variation (TV) regularizer [69]. This helps to better preserve discontinuities in the flow fields. Along with the use of the TV-regularizer, they extend an efficient projected gradient scheme proposed by Chambolle [17] that allows optical flow computation in real-time. Trobin et al. also considered the problem of piecewise constant flows in untextured regions [59]. Instead of applying a TV-L^1 regularization to the flow field, they propose to use an unbiased second-order regularizer to remove the bias towards constant flow fields. Werlberger et al. extended the isotropic Total Variation regularization approaches to an anisotropic regularization based on the robust Huber-norm [66]. Similar, Wedel et al. propose to use data-aware regularizers that adaptively favor rigid body motion if supported by the image data and motion field discontinuities that coincide with discontinuities of the image structure [62].

While all of the approaches so far focused on optimizing the data fidelity term and the smoothness term separately, Zimmer et al. recently investigated the interplay of these two terms [70]. Inspired by an early model by Nagel and Enkelmann [40], which regularizes the flow field along

edges but not across, they develop a synergistic model where data term and smoothness term do not interfere but complement each other instead. Their anisotropic smoothness term reduces smoothing in the data-constraint direction, while enforcing a strong filling-in effect orthogonal to it.

A different approach to tackle occluded and texture-less regions is taken by Xu et al. [68]. In their work, they infer a segment-based affine motion model using a color segmentation and an initial flow field obtained by Brox's method. After having inferred an occlusion map and a confidence map from the segments, the parametric motion model is incorporated in a variational framework, and the initial flow field is refined. Zitnick et al. also jointly estimate optical flow and segmentation and further allow fractional contributions of overlapping segments to individual pixels [71]. Based on the Gestalt principles of grouping, Werlberger et al. recently also incorporated a low level image segmentation into flow estimation [65].

Efficient minimization strategies. The minimization of the variational approaches is usually carried out in an iterative fashion, applying for example a projected gradient scheme. A drawback of such local iterative minimization techniques is their slow convergence and the risk of getting stuck in local minima. To account for this, Trobin et al. propose to approximate the minimization by solving a series of binary subproblems to facilitate large optimization moves [58]. Their proposed method can be interpreted as an extension of discrete graph-cut based methods such as α-expansion [13] or LogCut [31] to a spatially continuous setting.

3.4 Long-range methods

Despite the wide-spread use of multi-resolution methods in optical flow estimation, there are still cases where the displacement is too large to be estimated in a hierarchical framework. This is especially true if the object motion is larger than the object itself. Multi-resolution methods will not help in this case, since the object will vanish in the image pyramid before the displacement is small enough to be estimated. To account for this, Brox et al. recently incorporated descriptor matching in a variational framework to guide optical flow estimation for larger motions [14]. As descriptors, they propose to use regions descriptors of a hierarchical segmentation of the image, similar to the SIFT descriptor [34]. Their approach combines the power of descriptor matching with the regularization properties of a variational approach.

A different approach has been taken by Steinbrücker et al. [49]. Starting at a standard variational formulation and making use of techniques known from quadratic relaxation, they arrive at a formulation with a point-wise data term and a convex smoothness term which are coupled via an additional flow field. For both data and smoothness term, a globally optimal solution

can be found. The solution for the data term can simply be computed by a complete search, alleviating the need for coarse-to-fine warping strategies. Another appealing property of this approach is that any point-wise error measure can be integrated into the data term. This has been exploited in [50] where they integrate patch-based error measures into this framework.

While both approaches clearly improve on the current state-of-the-art for long range motions, they both suffer from a lower overall accuracy compared to warping-based methods [14, 49].

3.5 Discrete optimization

While variational methods seek to minimize the energy functional in a continuous domain, there are also approaches to optical flow computation using methods from discrete optimization. A common approach to optical flow estimation in a discrete setting is to formulate the process as maximum a-posteriori (MAP) inference. Contrary to continuous approaches, the solution is no longer continuous-valued but requires a sampling of the solution space. To this end, discrete methods usually have to deal accuracy for a computationally tractable label space.

Glocker et al. overlay the image with a uniform grid of control points and iteratively estimate displacement vectors for each control point based on Markov Random Fields and a warping strategy [24]. To account for the limited precision of a discrete label space, they estimate the uncertainty of the flow field in each control point and derive a new label space for every control point in every iteration. The dense optical flow field is obtained via cubic B-spline interpolation of the control points.

Lempitsky et al. combine discrete and continuous flow estimation [30]. Their algorithm fuses multiple proposal optical flow fields obtained using continuous optical flow estimation algorithms such as Horn-Schunck and Lucas-Kanade. Using a graph-cut optimization to decide for each pixel which proposal flow field the flow vector should be taken from, they achieve lower energy values than with either a pure discrete or purely continuous optimization approach.

3.6 Learning-based methods

Despite the long history of optical flow computation and the study of various data and smoothness terms, very few attempts have been taken to optimize those terms from actual data. In one of the first approaches to supervised learning optical flow [43], a detailed analysis of flow statistics in natural scenes is presented, and machine learning methods are developed to learn a Markov random field model of optical flow. The prior probability of a flow field is formulated as a Field-of-Experts model that captures the spatial statistics in overlapping patches and is trained using contrastive divergence.

This model is extended in Ref. [56] to the spatio-temporal domain to model temporal changes in image features. Further on, in this approach the statistical relationship between image and flow boundaries are modeled explicitly by a Steerable Random Field following the model proposed by Nagel and Enkelmann [40].

Instead of considering maximum likelihood estimation, Li and Huttenlocher [32] learn the parameters of a continuous-state Markov random field by minimizing the training loss for a set of ground-truth images. They use a technique from stochastic optimization, called simultaneous perturbation stochastic approximation, to optimize the error criterion used to evaluate the quality of the flow field. Their approach does not require approximations common in maximum-likelihood estimation and generalizes well to unseen data.

3.7 Occlusion handling

Occlusion handling is an important aspect of optical flow computation since no sensible correspondences can be found for occluded regions. If disregarded, the flow fields along occluding boundaries tend to collapse due to regularization. A first step towards accounting for occlusion was taken by jointly estimating forward and backward optical flow fields [2]. By jointly estimating both flow directions, occluded regions can be identified by examining the mismatch of forward and backward flow. Another approach to occlusion detection is to compute the divergence of the flow field and looking for areas with negative divergence. Sand and Teller [45] combine this approach with pixel projection differences to detect occluded regions and integrate this into the variational method by Brox et al. [15]. In their approach, they alternate optical flow estimation and occlusion detection. Similarly, Xiao et al. alternate optical flow computation and occlusion detection based on intensity mismatch [67]. Occluded regions are filled by adaptive bilateral filtering of the flow fields. In contrast to this, Ince and Konrad simultaneously estimate optical flow and occluded regions in a variational framework [28]. Optical flow in occluded regions is inpainted from neighboring visible regions using image-driven anisotropic diffusion.

A different approach is taken by Sellent et al. [48]: they propose an image formation model that relates a long-exposure image to preceding and succeeding short-exposure images in terms of optical flow and occlusion. With this method, not only binary occlusion maps but also the per-pixel occlusion time can be recovered. While originally also a two-step process, it has been recently integrated in a variational framework that allows simultaneous estimation of flow and occlusion information [47].

3.8 Performance evaluation

With a huge number of different optical flow algorithms available, datasets for qualitative evaluation of those algorithms have become an essential part of research. Starting with the benchmark set introduced by Barron et al. [8], the accuracy of optical flow algorithms has rapidly increased. However, after 13 years of research, the performance improvements on those benchmarks have largely saturated. To this end, Baker et al. recently established a new publicly available database [7] which focuses on current aspects of optical flow research, i.e. photo-realistic scenes with all artifacts of real sensors (noise, motion blur, etc.). This database has become a standard by now and it ranks the currently best optical flow algorithms with respect to different error measures. Nevertheless, slight criticism on the design of the test scenes recently arose in the computer vision community. Vaudrey et al. have shown in their report that high ranking algorithms often fail for real-world scenes, e.g. driving scenes with changing illumination and large motions [60].

In the following sections, we evaluate current optical flow research in the specific context of image interpolation, focusing on typical real-world scenes.

4 Optical Flow in Computer Graphics

In recent years, image-based rendering techniques have advanced from static to time-varying scenes. Instead of still images, multiple, not necessarily synchronized video streams now capture a dynamic scene from different viewpoints. This has lead to several new algorithms which try to extend findings for single images or pairs of images to video sequences. Despite the high processing power in modern PCs, however, it is often unavoidable to propagate some information along the sequence instead of recomputing it every frame in order to keep algorithms efficient. In addition, temporal coherence is an important aspect for visual fidelity of video streams. This section gives an overview of recent applications of optical flow in a computer graphics context.

Image and Video Registration. Image and video registration is the most prominent application of optical flow techniques in contemporary computer graphics. Applications range from image stabilization for hand-held acquisition [29, 57] and registration of video frames with different exposures [20, 64, 19] to registration of projected textures on the surface of an approximate 3D model [21]. Optical flow has also been used to register different video streams [44] as a preprocessing step for video editing. Most of these approaches rely on a gradient-based variant of the Lucas-Kanade algorithm [35]. As exceptions, Eisemann et al. [21] use the approach by Brox et al. [15], and Einarsson et al. [19] resort to the approach of Black and Anandan [11].

Information Propagation. Another important field of application for optical flow is information propagation along or between video streams. The most prominent application in this area is video matting, where information propagation is used to minimize tedious user interaction [18, 22, 5]. Depending on the required accuracy of the flow fields, authors resort to simple block-matching [22], local flow averaging [5] or the approach of Black and Anandan [18]. Einarsson et al. use flow fields to propagate captured reflectance fields along and between cameras for re-lighting purposes [19]. Peers et al. follows a similar approach to transfer reflectance fields for facial re-lighting [42]. Since both approaches require accurate correspondence fields, they use the algorithms by Black and Anandan [11] and Brox et al. [15], respectively.

Reconstruction and Augmentation. Dense correspondence fields are also needed by reconstruction or augmentation algorithms. Atcheson et al. use the Lucas-Kanade optical flow to extract the distortion in a high-frequency pattern introduced by a heated gas volume [4]. The 2D motion vectors then serve as a basis for the reconstruction of a refractive index field within a volume in a multi-camera setup. Scholz and Magnor [46] also use Lucas-Kanade flow to measure textile motion in a multi-camera setup and to reconstruct the 3D scene flow, which then serves as a basis for the animation of a virtual cloth. Hilsmann and Eisert use optical flow computed on a coarse mesh overlaid on the images to track textile motion in monocular sequences and to augment parts of the textiles with different textures [25, 26].

Image Interpolation. Concerning image interpolation, there are surprisingly few occurrences of traditional optical flow algorithms. Wang et al. use optical flow to warp images of asynchronously captured light fields to a common virtual time before reconstructing the virtual view [61]. Recently, Mahajan et al. proposed a path framework for image interpolation [36]. While the path framework does not compute optical flow fields in a traditional sense, they show that the paths can be transferred into the traditional optical flow representation. Stich et al. proposed an algorithm for deriving dense correspondence fields from sparse edge matches [53, 52] which serve as input for a perceptual image interpolation algorithm.

In the following section, we evaluate the current state-of-the-art in optical flow research for the task of image interpolation.

5 Optical Flow for Image Interpolation - A Case Study

In our case study, we systematically evaluate four different optical flow algorithms. The focus of this study is on the adequacy of the resulting flow

fields for use in multi-image interpolation algorithm as exployed for example in the Virtual Video Camera system **?**]. While optical flow research usually focuses on estimating correspondences from one frame of a camera to the next - in what follows referred to as *temporal image pair* - we are also interested in correspondence fields between frames captured by different cameras, referred to as *spatial image pairs*, and between frames captured by different cameras as well as different points in time, referred to as *spatio-temporal image pairs*. We thus evaluate each algorithm on three image pairs of our test dataset: a temporal pair, a spatial pair and a spatio-temporal pair. Each algorithm has to face several challenges common to correspondence estimation on space-time navigation footage:

1. large pixel distances in the order of up to 20% of the image diagonal, especially for spatial and spatio-temporal image pairs
2. fast moving objects / small objects
3. changing illumination for spatial and spatio-temporal image pairs
4. occlusions and disocclusion due to viewpoint change and motion
5. large untextured regions

5.1 Selection of optical flow algorithms

To evaluate the state-of-the-art in optical flow computation, we deliberately chose the two published top performers on Middlebury's flow evaluation data base [7] at the time of this writing.

The algorithm by Sun et al. [55] performs best with respect to the endpoint error, which is considered to be the gold standard for evaluating optical flow accuracy. Sun et al. thoroughly investigated how the objective function, the optimization method, and modern implementation practices influence flow accuracy. They combine their findings with a weighted non-local median filtering term in the classical Horn-Schunck model. The weighted non-local median filtering approach avoids over-smoothing fine image details. With respect to the normalized interpolation error, this algorithms is ranked 19th out of forty.

The algorithm by Werlberger et al. [66] performs best with respect to the normalized interpolation error.This error measure most closely reflects the goal of this study, i.e. interpolation quality. In their approach, Werlberger et al. increase the robustness as well as the accuracy of discontinuity preserving variational optical flow models by replacing the isotropic total variation regularization with an image-driven anisotropic one based on the robust Huber-L^1-norm. They further propose to exploit symmetry around a central frame if more than two images are available.

In addition to the two top performers described above, we also include the algorithm proposed by Steinbrücker et al. [49] in our study. This approach is especially suited for estimating fast, long-range motion without resorting to a coarse-to-fine warping strategy. While pyramidal approaches are able to handle large motions in principle, they still fail as soon as the displacement of the object is larger than the extent of the object itself. This is a well-known limitation of optical flow estimation and has recently been tackled by several researchers independently [14, 49]. We favor the approach by Steinbrücker et al. for its flexibility in the choice of the point-wise data term. This allows to integrate arbitrary point-wise descriptors, e.g. SIFT descriptors [34], into the flow estimation. However, we stick with the original formulation [49] for our comparison.

Finally, we evaluate our test scenes on correspondence fields computed by the algorithm proposed by Stich et al. [52]. The main focus of Stich et al. lies on correctly matching edges and moving regions coherently, inspired by findings from perceptual research. To this end, Stich et al. propose to match edge pixels, followed by a least-squares estimation of a perspective transformation for each image region based on the matches and an initial super-segmentation of the image. Their approach can be considered a piecewise constant optical flow. While this algorithm also offers the possibility for manual correction of flow fields, we used uncorrected flow fields for fair comparison.

Except for the algorithm by Steinbrücker et al., we used publicly available implementations of the authors. All algorithms except the one by Werlberger et al. operate on color images. For this algorithm, we had to desaturate the images first. The parameters for the different algorithms were set to default values (if provided in the original references) or optimized for the spatio-temporal pair and kept fix for the spatial and temporal pairs. While one can usually achieve better results by tuning the parameters for each image pair individually, we opt for this approach since we want to be able to process thousands of image pairs, rendering individual parameter tuning impossible.

5.2 Interpolation method

We evaluate the performance of the selected algorithms by forward-warping both images and adaptively blending them to obtain the interpolated image [52]. The blending weights are determined per pixel from the connectedness of the motion fields [37]. Lacking depth information, Stich et al. propose to use simple heuristics to infer a relative depth ordering from the motion fields. For convenience, we summarize the algorithm in Alg. 1. While there are other interpolation algorithms that are based on backward warping, e.g. the one proposed by Baker et al. [6] and used in the Middlebury evaluation, we stick with the forward-warping approach because it allows

Input: frames I_0, I_1, flow fields u_0, u_1, t
Output: interpolated frame I_t
begin
`// forward warp` I_0
$I'_t(\text{round}(x + t \cdot u_0(x))) = I_0(x)$;
`// forward warp` I_1
$I'_{1-t}(\text{round}(x + (1-t) \cdot u_1(x))) = I_1(x)$;
compute connectedness c_{u_0}, c_{u_1};
`// adaptively blend based on connectedness` c_{u_0} `and` c_{u_1}
$I_t(x) = t \cdot c_{u_0}(x) \cdot I'_t(x) + (1-t) \cdot c_{u_1}(x) \cdot I'_{1-t}(x)$;
end

Algorithm 1: Perceptual interpolation algorithm proposed by Stich et al. [52]

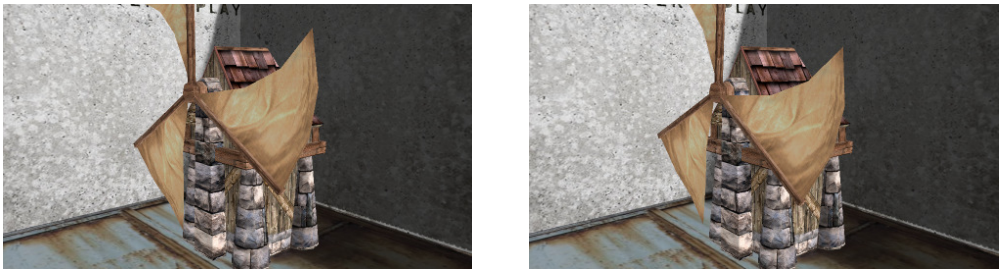
also for interpolation from multiple images needed later.

5.3 Test scenes

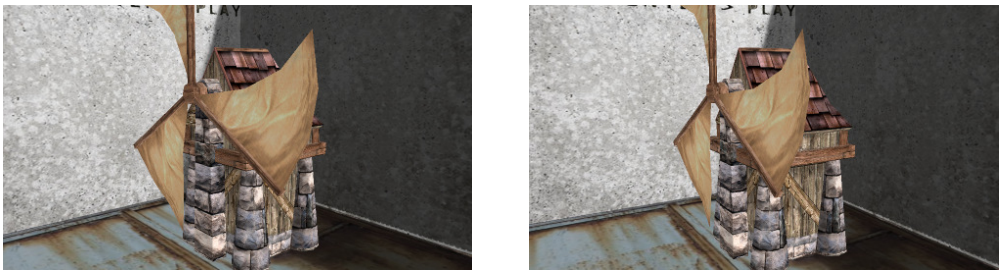
Our test set includes one synthetic test sequence with known ground truth motion fields and depth, and three real-world scenes of varying complexity and different challenges. For the real world data set, ground truth motion as well as ground truth interpolation results are not available but made by hand. The synthetic *Stonemill* sequence features long-range motion of up to 60 pixels at a resolution of 480×270 pixels and large occlusion/disocclusion areas, Fig. 1. The *Mona* sequence is a real-world sequence that was captured under controlled studio illumination, Fig. 2. It features large displacements of thin structures (arms, hat stands) as well as large untextured regions. The maximal pixel displacement is around 90 pixels at a resolution of 960×540 pixels. The *Skateboarder* sequence adds varying illumination conditions and shadows due to outdoor capture to the set of challenges, Fig. 3. The maximal pixel displacement for the tested image pairs is around 180 pixels at a resolution of 960×540 pixels. It further features large disoccluded regions at the boundary and behind the skateboarder. The *Parkour* sequence is the most complex scene in our test set, Fig. 4. The complexity arises from the background with a lot of occlusion/disocclusion, the fine structures of the twigs and varying illumination. The challenges along with the image resolution and maximal displacements are listed in Table 1.

5.4 Evaluation

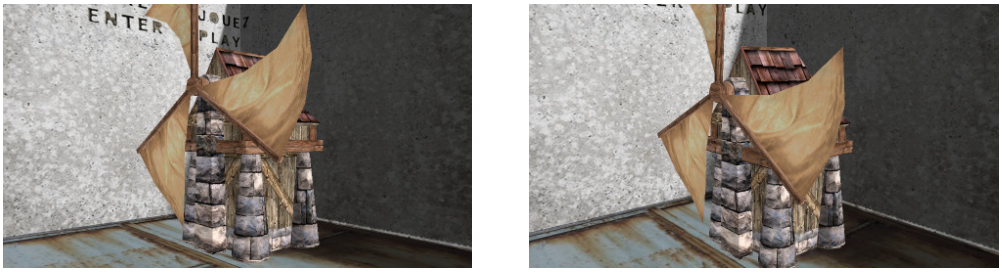
We compute correspondence fields for all image pairs of our test set using the four selected optical flow algorithms. In addition, we generate ground truth motion fields for the real-world scenes by hand. Ground truth in this case reads as "flow fields producing the visually most plausible interpolation



(a) Temporal pair



(b) Spatial pair



(c) Spatio-temporal pair

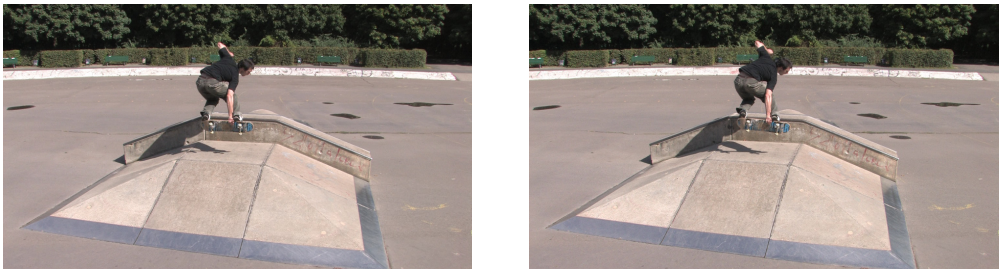
Figure 1: Image pairs used for evaluation on the Stonemill scene.

(a) Temporal pair

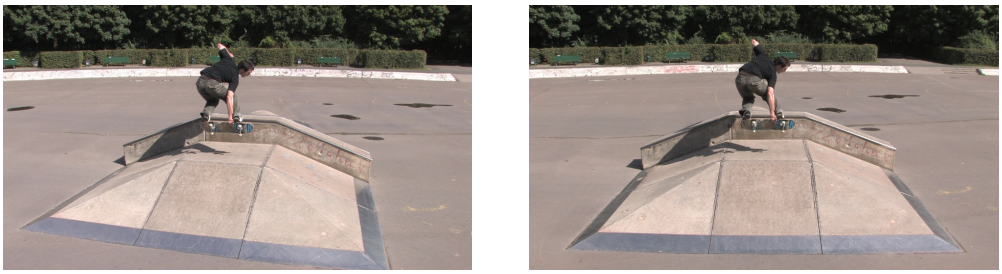
(b) Spatial pair

(c) Spatio-temporal pair

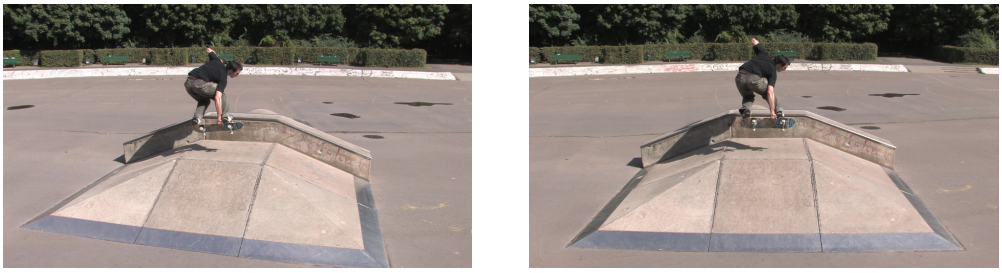
Figure 2: Image pairs used for evaluation on the Mona scene.



(a) Temporal pair



(b) Spatial pair



(c) Spatio-temporal pair

Figure 3: Image pairs used for evaluation on the Skateboarder scene.

(a) Temporal pair

(b) Spatial pair

(c) Spatio-temporal pair

Figure 4: Image pairs used for evaluation on the Parkour scene.

Scene	Resolution (pixel)	Max. Disp. (pixel)	Challenges
Stonemill	480×270	60	long-range motion, large occluded/disoccluded areas
Mona	960×540	90	untextured regions, fast motion of fine structures
Skateboarder	960×540	180	long-range motion, large occluded/disoccluded areas, varying illumination
Parkour	960×540	170	complex background, multiple motion layers, occlusion/disocclusion, varying illumination

Table 1: Test scenes used for evaluation along with their specific challenges.

sequence", not necessarily the physically correct motion vectors. We then interpolate an image sequence between the two source images using Alg. 1 and evaluate the quality. All scenes are evaluated in a perceptual user study including interpolation sequences generated from ground truth flow fields; the synthetic sequence is evaluated numerically as well using synthesized ground truth images. A numerical evaluation of the interpolation error on the real-world sequences would only be possible for the temporal image pairs since each camera provides enough frames along the temporal dimension for a leave-one-out comparison. The spatial and spatio-temporal pairs, however, are sampled too sparsely in the existing test footage; skipping a camera and/or frame for a leave-one-out comparison would result in too large pixel displacements and insufficient image overlap.

Synthetic sequence with ground truth

We first numerically evaluate the quality of the individual algorithms using the synthetic *Stonemill* sequence along with the ground truth correspondence fields. To this end, we compute the error measures *average endpoint error (AEE)*, *average angular error (AAE)*, *interpolation error (IE)* and *normalized interpolation error (NIE)* as proposed in Ref. [6]. The results are summarized in Table 2. Figs. 5, 6, 7 show the interpolation result along with a contrast-stretched visualization of the interpolation error. The numerical evaluation confirms the Middlebury ranking: the approach by Sun et al. [55] performs best, on average, with respect to the angular and endpoint error measures. The algorithm by Werlberger et al. [66] ranks highest with respect to the interpolation and normalized interpolation error. The long-range method by Steinbrücker et al. [49] cannot compete with the two

top-performers with respect to the numerical accuracy of the flow fields due to the discrete sampling of the motion vectors. The approach by Stich et al. [52] was designed with perceptually plausible interpolation results as goal. As such, it is expected to perform poorest with respect to the numerical evaluation.

The numerical analysis is restricted to a single frame in the middle of the interpolation sequence. Thus, it naturally masks artifacts arising due to motion along the sequence. However, temporal consistency is an important aspect in terms of interpolation quality. To capture this aspect, we evaluate the visual quality of the whole sequence with a psychophysical user study.

	Pair	Error	Stich[52]	Werlberger[66]	Sun[55]	Steinbrücker[49]
spatial	$\mathbf{w}_{1,2}$	AAE	57.64	7.56	3.86	11.02
		AEE	16.80	2.99	1.19	7.08
	$\mathbf{w}_{2,1}$	AAE	37.45	34.28	3.64	17.57
		AEE	11.44	8.63	1.64	8.80
		IE	40.45	28.57	29.29	34.08
		NIE	4.08	2.93	3.01	3.57
spatio-temporal	$\mathbf{w}_{1,2}$	AAE	42.37	19.64	22.19	11.75
		AEE	12.23	5.38	5.68	5.60
	$\mathbf{w}_{2,1}$	AAE	43.91	15.12	16.55	17.84
		AEE	14.07	4.97	4.86	7.44
		IE	33.45	28.10	29.60	33.30
		NIE	3.05	2.51	2.69	3.12
temporal	$\mathbf{w}_{1,2}$	AAE	12.45	2.31	0.78	4.00
		AEE	1.45	0.18	0.07	0.68
	$\mathbf{w}_{2,1}$	AAE	8.34	2.53	0.82	4.15
		AEE	0.74	0.21	0.06	0.68
		IE	11.26	11.07	11.47	8.19
		NIE	0.91	0.87	0.84	0.54

Table 2: Numerical evaluation of synthetic test sequence. Best values are marked in green.

User study

In order to assess the perceptual quality of the interpolation sequences for real-world scenes as well as the synthetic sequence, we followed the approach presented in Ref. [52] and carried out a user study. The two major goals of the study were

(1) to investigate whether any of the tested algorithms attains a perceptual quality score similar to ground truth, and

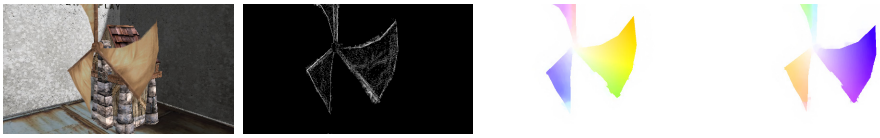
(2) to compare the results of the tested algorithms against each other and assess whether there is a statistical significance with respect to perceived quality.

Rather than using a standard rating task in which participants would be shown a sequence and be asked to rate its quality, we opted for a more systematic approach. In the psychophysical study, we used a two-alternative-forced-choice task in which two interpolation sequences were shown successively, and participants were asked to indicate which sequence contained less visual artifacts. Such a direct comparison allows for a more fine-grained analysis of the data as rating tasks are often subject to scaling problems. For each of the 4 different test scenes, we compared all 4 different optical flow algorithms as well as interpolations created from ground-truth flow fields against each other (only doing pairwise AB and AA, not BA comparisons), yielding a total of $4 \cdot 3 \cdot (4 \cdot \frac{5}{2} + 5) = 180$ trials. To meet the first goal of our study, we included hand-made correspondence fields as ground-truth into the user study.

All real-world scenes were rendered at a resolution of 960x540 pixels with 25 frames per second and were 2–4 seconds long. Figs. 11, 12, 13 show the interpolation result halfway between the image pairs. The synthetic scene was rendered at a resolution of 480x270 pixels. The sequences were presented on a black-background LCD monitor using a resolution of 1366x768 pixel at 60 Hz. Participants viewed the stimuli at a distance of roughly 50 cm. Each trial consisted of a fixation cross shown for 1 second, followed by the first sequence, a second fixation cross shown for 0.5 seconds, and the second sequence. After this, the screen was blanked and participants were asked to indicate by key press which sequence contained less visual artifacts. Participants were briefed before the experiment that artifacts were defined as “any visual disturbance resulting in non-smooth transition”. All participants completed two test trials before the experiment to familiarize them with the task. Neither during the test trials nor the experiment was any feedback given. The whole experiment lasted approximately 35 minutes. Our test group consisted of 9 participants who had *strong* computer graphics-related experience.

Figure 5: Interpolation results on the temporal stonemill pair. Rows 2–5 show (left to right) the interpolation result, a gamma-corrected visualization of the interpolation error, and forward and backward flow fields. The approach by Steinbrücker et al. [49] yields the lowest interpolation error, cf. Table 2.

(a) Interpolation result using ground truth flow fields. From left to right: ground truth $I_{1.5}$, ground truth flow fields $\mathbf{w}_{1,2}$ and $\mathbf{w}_{2,1}$.



(b) Interpolation result using flow fields generated with Sun et al. [55]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

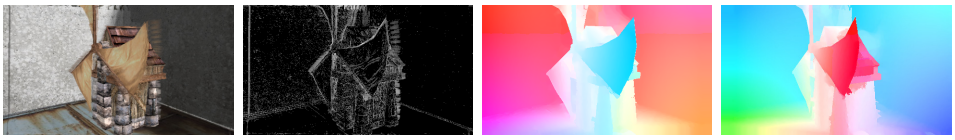
(c) Interpolation result using flow fields generated with Steinbrücker et al. [49]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

(d) Interpolation result using flow fields generated with Werlberger et al. [66]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

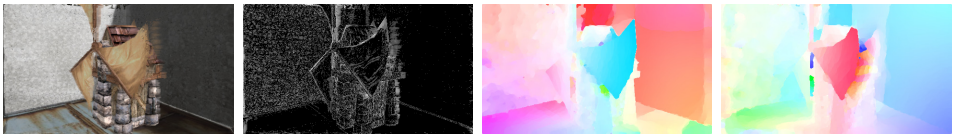
(e) Interpolation result using flow fields generated with Stich et al. [52]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

Figure 6: Interpolation results on the spatial stonemill pair. Rows 2–5 show (left to right) the interpolation result, a gamma-corrected visualization of the interpolation error, and forward and backward flow fields. The approach by Werlberger et al. [66] yields the lowest interpolation error, cf. Table 2.

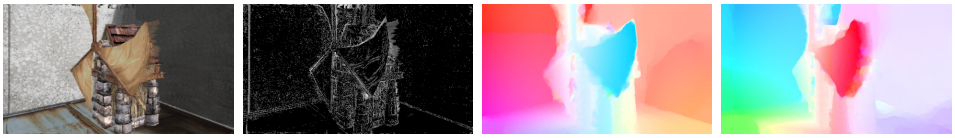
(a) Interpolation result using ground truth flow fields. From left to right: ground truth $I_{1.5}$, ground truth flow fields $\mathbf{w}_{1,2}$ and $\mathbf{w}_{2,1}$.



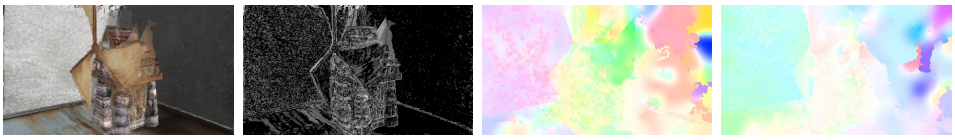
(b) Interpolation result using flow fields generated with Sun et al. [55]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.



(c) Interpolation result using flow fields generated with Steinbrücker et al. [49]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

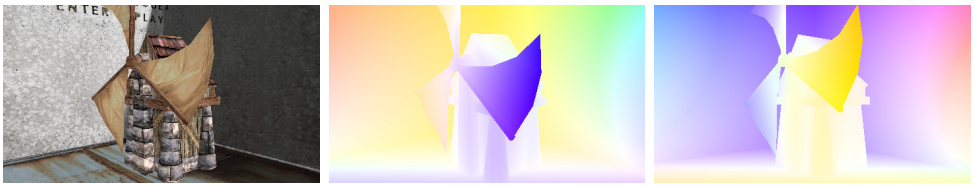


(d) Interpolation result using flow fields generated with Werlberger et al. [66]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

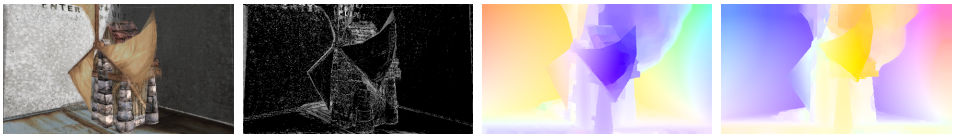


(e) Interpolation result using flow fields generated with Stich et al. [52]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

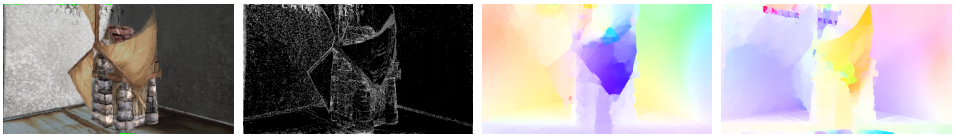
Figure 7: Interpolation results on the spatio-temporal stonemill pair. Rows 2–5 show (left to right) the interpolation result, a gamma-corrected visualization of the interpolation error, and forward and backward flow fields. The approach by Werlberger et al. [66] yields the lowest interpolation error, cf. Table 2.



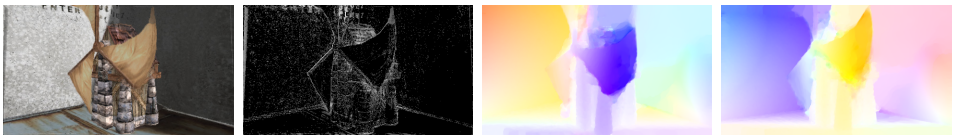
(a) Interpolation result using ground truth flow fields. From left to right: ground truth $I_{1.5}$, ground truth flow fields $\mathbf{w}_{1,2}$ and $\mathbf{w}_{2,1}$.



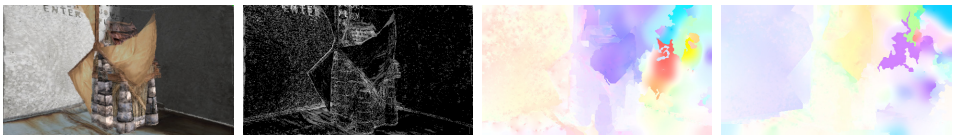
(b) Interpolation result using flow fields generated with Sun et al. [55]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.



(c) Interpolation result using flow fields generated with Steinbrücker et al. [49]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.



(d) Interpolation result using flow fields generated with Werlberger et al. [66]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.



(e) Interpolation result using flow fields generated with Stich et al. [52]. From left to right: $I_{1.5}$, interpolation error, $\mathbf{w}_{1,2}$, $\mathbf{w}_{2,1}$.

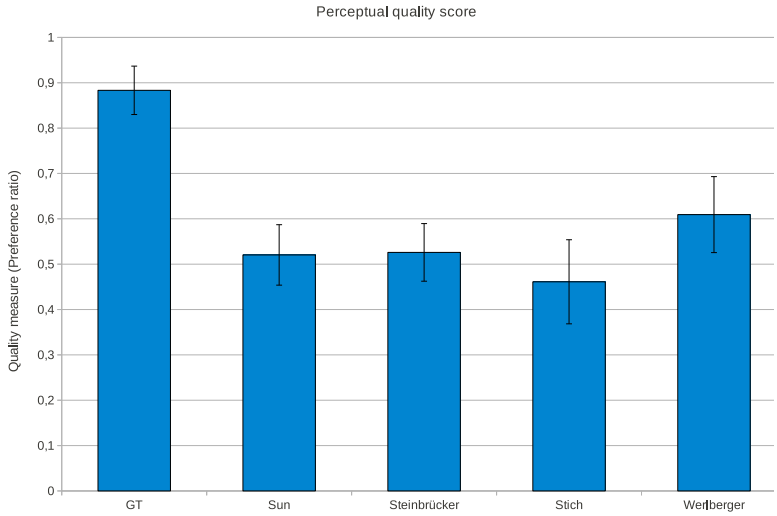


Figure 8: Perceptual quality scores for 5 different test conditions (optical flow algorithms).

Analysis

For the first analysis, we determined a perceptual quality score for each algorithm by counting how many times it was chosen as producing fewer visual artifacts when compared to one of the other algorithms. The normalized scores are shown in Fig. 8 for all five approaches. As can be seen immediately, there is a significant difference between perceived quality of interpolations based on ground-truth flow fields and perceived quality of all interpolations generated from the other algorithms. This is confirmed by the highly significant one-way anova ($F(4, 40) = 46.546, p = 0.001$). A Tukey test [12] with confidence level $p = 0.01$ further reveals that only the ground truth interpolations differ significantly.

In order to answer the first question of our study, we re-plot the data in Fig. 8 to show how often participants would choose any other algorithm over ground truth interpolation, that is, how many times the perceptual quality of the sequence was at least as good as in the ground truth case. We break this analysis down by test scenes. First of all, Fig. 9 confirms the results outlined above: none of the four optical flow methods are selected more than 1 or 2 times, on average (out of a maximum of 9), indicated by the colored dashed lines in Fig. 9, over the ground truth interpolation for all scenes.

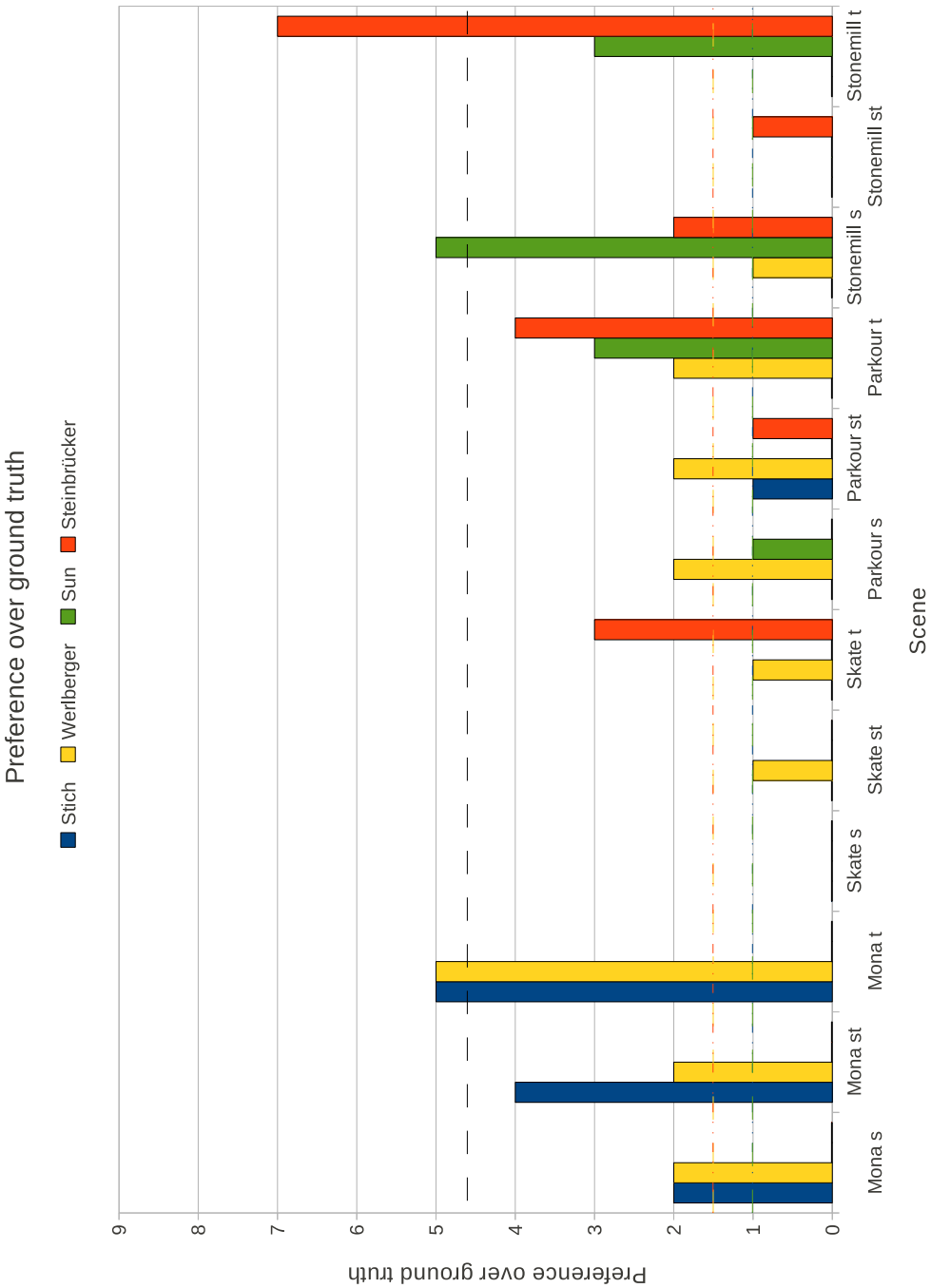


Figure 9: Preference of optical flow algorithms over ground truth flow fields, broken down by test scene. Values around 4.5 denote that both conditions are of equal perceived quality. None of the tested algorithms is able to reach this goal.

Values around 4.5 in Fig. 9 would indicate equal perceived quality. Clearly, none of the tested algorithms can reach this threshold consistently over all tested scenes.

However, with respect to the individual scenes, there are noticeable differences. With respect to the *Mona* scene, the algorithm by Stich et al. [52] provides a perceptual quality comparable to ground truth; the algorithm by Werlberger et al. [66] fares only slightly worse. This scene is extremely well suited for the approach of Stich et al. since most errors are hidden in the large untextured regions in the background, and the important edge structures are distinct and can be matched unambiguously. Also, the algorithm by Werlberger et al. can score on this scene due to its powerful anisotropic regularization. The approach by Sun et al. [55] fails to recover the motion of the arm and also diffuses wrong motion information into background, leading to noticeable artifacts. While the interpolation created from Steinbrücker et al.'s flow fields is able to maintain all fine structures in the background, the visual quality is heavily impaired by wrong matches in the untextured regions between the legs and in the disoccluded region behind the actor's head. Furthermore, the regularization here also diffuses wrong motion information into the background.
Considering the more complex *Skateboarder* scene, no algorithm is able to achieve ground truth quality. For the edge-based algorithm by Stich et al., this scene already exhibits a too complex scene structure where the edge matching approach clearly fails. The top-ranking algorithms on Middlebury provide a good interpolation for the foreground object. However, the background gets distorted, and the puddles on the ground move in an unnatural way. Steinbrücker et al.'s algorithm is the only one that manages to transform visible parts of the background correctly, but again suffers from spurious wrong matches that ruin overall interpolation quality.
Surprisingly, for the most complex *Parkour* scene, every algorithm can collect some votes. This might be due to the fact that even the ground truth interpolation suffers from artifacts arising from the depth heuristics employed in the interpolation algorithm. However, only the algorithm by Steinbrücker et al. [49] reaches a perceptual quality comparable to ground truth for the temporal interpolation.
For the synthetic test scene, the approach by Sun et al. [55] is on par with ground truth for the spatial image pair, whereas the algorithm by Steinbrücker et al. outperforms ground truth for the temporal pair. However, both algorithms are not able to perform equally well on the other two image pairs. The algorithms by Werlberger et al. and Stich et al. fail to maintain important structures of the building.

To address the second goal of our study, we repeat the evaluation without taking the comparisons against ground truth into account, Fig. 10. The insignificant one-way anova ($F(3, 32) = 5.354, p = 0.001$) indicates that there is no statistical difference between the individual approaches with respect

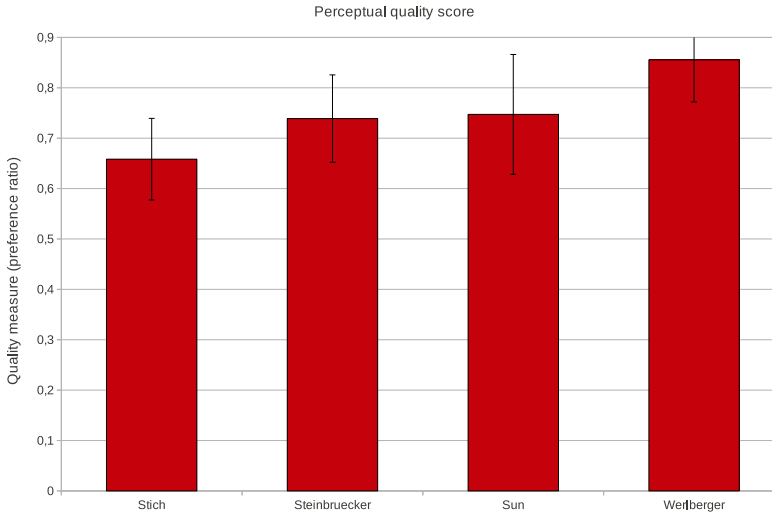


Figure 10: Perceptual quality scores leaving out comparisons against ground truth.

to interpolation quality. Again, a Tukey test with significance level $p = 0.01$ confirms this.

6 Conclusion

To summarize, we have evaluated four recent optical flow algorithms in the context of image interpolation. We evaluated their perceptual quality on three typical real-world scenes used in the Virtual Video Camera system [33]. A psychophysical user study has shown that no algorithm consistently reaches ground truth interpolation quality on every scene. The approach proposed by Stich et al. [52] performs well on scenes with a simple background and a few distinct edges, but fails as soon as the edge structure of the images becomes too complex. The current top-performer on Middlebury with respect to the angular and endpoint error measures [55] only produces convincing results on the synthetic sequence, but produces noticeable artifacts on all real-world scenes. The approach by Werlberger et al. [66] produces good results on scenes with moderate motion and is at its most impressive in untextured regions. Similar to Stich et al. [52], it maintains important edge structures due to the anisotropic regularization. However, being based on a pyramidal approach, it fails to recover the motion of small, fast objects. In contrast to this, the long-range method by Steinbrücker et al. [49] is able to recover fast motion of small objects, even over large pixel distance. Unfortunately, it suffers from spurious wrong matches resulting from ambiguities in the global optimization approach. This most

Figure 11: Interpolation results on the temporal image pairs. For the Mona sequence (first column), the algorithms by Stich et al. [52] (fifth row) and Werlberger et al. [66] (fourth row) are on par with ground truth (first row). For the Skateboarder sequence, no algorithm is able to reach ground truth interpolation quality. For the Parkour sequence (third column), only the approach by Steinbrücker et al. [49] (third row) obtains a quality comparable to ground truth.

(a) Interpolation results on the temporal pairs using ground truth flow fields.

(b) Interpolation results on the temporal pairs using flow fields generated with Sun et al. [55].

(c) Interpolation results on the temporal pairs using flow fields generated with Steinbrücker et al. [49].

(d) Interpolation results on the temporal pairs using flow fields generated with Werlberger et al. [66].

(e) Interpolation results on the temporal pairs using flow fields generated with Stich et al. [52].

Figure 12: Interpolation results on the spatial image pairs. No algorithm reaches ground truth interpolation quality for any sequence.

(a) Interpolation results on the spatial pairs using ground truth flow fields.

(b) Interpolation results on the spatial pairs using flow fields generated with Sun et al. [55].

(c) Interpolation results on the spatial pairs using flow fields generated with Steinbrücker et al. [49].

(d) Interpolation results on the spatial pairs using flow fields generated with Werlberger et al. [66].

(e) Interpolation results on the spatial pairs using flow fields generated with Stich et al. [52].

Figure 13: Interpolation results on the spatio-temporal image pairs. Only on the Mona sequence (first column), the algorithm by Stich et al. [52] (fifth row) is able to reach a quality comparable to ground truth. On the Skateboarder and Parkour sequences, no algorithm reaches this goal.

(a) Interpolation results on the spatio-temporal pairs using ground truth flow fields.

(b) Interpolation results on the spatio-temporal pairs using flow fields generated with Sun et al. [55].

(c) Interpolation results on the spatio-temporal pairs using flow fields generated with Steinbrücker et al. [49].

(d) Interpolation results on the spatio-temporal pairs using flow fields generated with Werlberger et al. [66].

(e) Interpolation results on the spatio-temporal pairs using flow fields generated with Stich et al. [52].

strongly shows up in disoccluded areas where all of the tested algorithms have problems.

As a conclusion, no algorithm is directly applicable for multi-view interpolation. The optimal algorithm would be a combination of a strong anisotropic regularization as proposed by Werlberger et al. [66] with a long-range, global optimization approach as proposed by Steinbrücker et al. [49].

References

[1] Aggarwal, J. and Nandhakumar, N. 1988. On the computation of motion from sequences of images: A review. In *Proceedings of the IEEE*. Vol. 76. 917–935.

[2] Alvarez, L., Deriche, R., Papadopoulo, T., and Sanchez, J. 2002. Symmetrical Dense Optical Flow Estimation with Occlusion Detection. In *Proc. of European Conference on Computer Vision (ECCV'02)*. Springer, 721–735.

[3] Atcheson, B., Heidrich, W., and Ihrke, I. 2009. An evaluation of optical flow algorithms for background oriented schlieren imaging. *Experiments in Fluids 46,* 3, 467–476.

[4] Atcheson, B., Ihrke, I., Heidrich, W., Tevs, A., Bradley, D., Magnor, M., and Seidel, H.-P. 2008. Time-resolved 3d capture of non-stationary gas flows. In *Proc. of ACM SIGGRAPH Asia 2008*. ACM, New York, NY, USA, 1–9.

[5] Bai, X., Wang, J., Simons, D., and Sapiro, G. 2009. Video SnapCut: robust video object cutout using localized classifiers. *ACM Trans. on Graphics 28,* 3, 70:1–70:11.

[6] Baker, S., Roth, S., Scharstein, D., Black, M. J., Lewis, J., and Szeliski, R. 2007. A database and evaluation methodology for optical flow. In *IEEE International Conference on Computer Vision (ICCV'07)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–8.

[7] Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., and Szeliski, R. 2010. Middlebury Optical Flow Evaluation Database. http://vision.middlebury.edu/flow.

[8] Barron, J., Fleet, D., and Beauchemin, S. 1994. Performance of optical flow techniques. *IJCV 12,* 1, 43–77.

[9] Battiti, R., Amaldi, E., and Koch, C. 1991. Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy. *IJCV 6,* 2 (June), 133–145.

[10] Beauchemin, S. S. and Barron, J. L. 1995. The computation of optical flow. *ACM Comput. Surv. 27,* 3, 433–466.

[11] Black, M. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields. *Computer Vision and Image Understanding 63,* 1 (January), 75–104.

[12] Box, G. E., Hunter, J. S., and Hunter, W. G. 2005. *Statistics for experimenters: Design, Innovation, and Discovery.* Wiley-Interscience; 2 edition.

[13] Boykov, Y., Veksler, O., and Zabih, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell. 23,* 11, 1222–1239.

[14] Brox, T., Bregler, C., and Malik, J. 2009. Large Displacement Optical Flow. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09).* to appear.

[15] Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proc. of European Conference on Computer Vision (ECCV'04).* Vol. 4. 25–36.

[16] Bruhn, A., Weickert, J., and Schnörr, C. 2005. Lucas/kanade meets horn/schunck: combining local and global optic flow methods. *IJCV 61,* 3, 211–231.

[17] Chambolle, A. 2004. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis. 20,* 1-2, 89–97.

[18] Chuang, Y.-Y., Agarwala, A., Curless, B., Salesin, D. H., and Szeliski, R. 2002. Video matting of complex scenes. *ACM Trans. on Graphics 21,* 3 (July), 243–248. Special Issue of the SIGGRAPH 2002 Proceedings.

[19] Einarsson, P., Chabert, C.-F., Jones, A., Lamond, B., Ma, W.-C., Sylwan, S., Hawkins, T., and Debevec, P. 2006. Relighting Human Locomotion with Flowed Reflectance Fields. In *Proc. of Eurographics Rendering Workshop.* Eurographics Association, 183–194.

[20] Eisemann, E. and Durand, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Trans. on Graphics 23,* 3, 673–678.

[21] Eisemann, M., Decker, B. D., Magnor, M., Bekaert, P., de Aguiar, E., Ahmed, N., Theobalt, C., and Sellent, A. 2008. Floating Textures. *Computer Graphics Forum 27,* 2 (4), 409–418.

[22] Eisemann, M., Wolf, J., and Magnor, M. 2009. Spectral Video Matting. In *Proc. of Vision, Modeling and Visualization (VMV 2009).* Braunschweig, Germany, 121–126.

[23] Enkelmann, W. 1988. Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics, and Image Processing 43,* 2 (August), 150–177.

[24] Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., and Navab, N. 2008. Optical flow estimation with uncertainties through dynamic mrfs. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'08)*. 1–8.

[25] Hilsmann, A. and Eisert, P. 2008. Optical flow based tracking and retexturing of garments. In *Proc. of IEEE International Conference on Image Processing (ICIP'08)*. San Diego, California, USA.

[26] Hilsmann, A. and Eisert, P. 2009. Realistic cloth augmentation in single view video. In *Proc. of Vision, Modeling and Visualization (VMV 2009)*. Braunschweig, Germany.

[27] Horn, B. K. P. and Schunck, B. G. 1981. Determining optical flow. *Artificial Intelligence 17*, 185–203.

[28] Ince, S. and Konrad, J. 2008. Occlusion-Aware Optical Flow Estimation. *IEEE Trans. Image Processing 17,* 8, 1443–1451.

[29] Kang, S., Uttendaele, M., Winder, S., and Szeliski, R. 2003. High Dynamic Range Video. *ACM Trans. on Graphics 22,* 3, 319–325.

[30] Lempitsky, V., Roth, S., and Rother, C. 2008. Fusionflow: Discrete-continuous optimization for optical flow estimation. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'08)*. 1–8.

[31] Lempitsky, V., Rother, C., and Blake, A. 2007. Logcut - efficient graph cut optimization for markov random fields. In *IEEE International Conference on Computer Vision (ICCV'07)*. 1 –8.

[32] Li, Y. and Huttenlocher, D. P. 2008. Learning for optical flow using stochastic optimization. In *Proc. of European Conference on Computer Vision (ECCV'08)*. Springer-Verlag, Berlin, Heidelberg, 379–391.

[33] Lipski, C., Linz, C., Berger, K., Sellent, A., and Magnor, M. 2010. Virtual video camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum 29,* 8 (Dezember), 2555–2568.

[34] Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV 60,* 2, 91–110.

[35] Lucas, B. and Kanade, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proc. of International Joint Conference on Artificial Intelligence*. Vancouver, Canada, 674–679.

[36] Mahajan, D., Huang, F., Matusik, W., Ramamoorthi, R., and Belhumeur, P. 2009. Moving Gradients: A Path-Based Method for

Plausible Image Interpolation. *ACM Trans. on Graphics 28,* 3, 42:1–42:11.

[37] Mark, W., McMillan, L., and Bishop, G. 1997. Post-Rendering 3D Warping. In *Proc. of Symposium on Interactive 3D Graphics (I3D'97)*. 7–16.

[38] McCane, B., Novins, K., Crannitch, D., and Galvin, B. 2001. On benchmarking optical flow. *Computer Vision and Image Understanding 84*, 126–143.

[39] Mitiche, A. and Bouthemy, P. 1996. Computation and analysis of image motion: A synopsis of current problems of methods. *IJCV 19,* 1, 29–55.

[40] Nagel, H. H. and Enkelmann, W. 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell. 8,* 5, 565–593.

[41] Otte, M. and Nagel, H.-H. 1994. Optical flow estimation: advances and comparisons. In *Proc. of European Conference on Computer Vision (ECCV'94)*. 51–60.

[42] Peers, P., Tamura, N., Matusik, W., and Debevec, P. 2007. Post-production Facial Performance Relighting using Reflectance Transfer. *ACM Trans. on Graphics 26,* 3, 52–62.

[43] Roth, S. and Black, M. J. 2007. On the Spatial Statistics of Optical Flow. *IJCV 74,* 1, 33–50.

[44] Sand, P. and Teller, S. 2004. Video matching. *ACM Trans. on Graphics 23,* 3, 592–599.

[45] Sand, P. and Teller, S. 2008. Particle Video: Long-Range Motion Estimation Using Point Trajectories. *IJCV 80,* 1 (October), 72–91.

[46] Scholz, V. and Magnor, M. 2004. Cloth motion from optical flow. In *Proc. of Vision, Modeling, and Visualization (VMV 2004)*. 117–123.

[47] Sellent, A., Eisemann, M., Goldlücke, B., Pock, T., Cremers, D., and Magnor, M. 2009. Variational Optical FLow from Alternate Exposure Images. In *Proc. of Vision, Modeling and Visualization (VMV 2009)*.

[48] Sellent, A., Eisemann, M., and Magnor, M. 2009. Motion Field and Occlusion Time Estimation via Alternate Exposure Flow. In *Proc. of IEEE International Conference on Computational Photography (ICCP'09)*. IEEE.

[49] Steinbrücker, F., Pock, T., and Cremers, D. 2009. Large Displacement Optical Flow Computation without Warping. In *IEEE International Conference on Computer Vision (ICCV'09)*. Kyoto, Japan.

[50] Steinbrücker, F., T.Pock, and D.Cremers. 2009. Advanced Data Terms for Variational Optic Flow Estimation. In *Proc. of Vision, Modeling and Visualization (VMV 2009)*. 155–162.

[51] Stich, T., Linz, C., Albuquerque, G., and Magnor, M. 2008. View and Time Interpolation in Image Space. *Computer Graphics Forum 27,* 7, 1781–1787.

[52] Stich, T., Linz, C., Wallraven, C., Cunningham, D., and Magnor, M. 2010. Perception-motivated Interpolation of Image Sequences. *ACM Transactions on Applied Perception (TAP)*, to appear.

[53] Stich, T., Linz, C., Wallraven, C., Cunningham, D., and Magnorw, M. 2008. Perception-motivated Interpolation of Image Sequences. In *Proc. of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*. ACM, Los Angeles, USA.

[54] Stiller, C. and Konrad, J. 1999. Estimating motion in image sequences: A tutorial on modeling and computation of 2d motion. *IEEE Signal Processing Magazine 16,* 4, 70–91.

[55] Sun, D., Roth, S., and Black, M. J. 2010. Secrets of optical flow estimation and their principles. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'10)*. IEEE Computer Society, 2432–2439.

[56] Sun, D., Roth, S., Lewis, J. P., and Black, M. J. 2008. Learning optical flow. In *Proc. of European Conference on Computer Vision (ECCV'08)*. Springer-Verlag, Berlin, Heidelberg, 83–97.

[57] Telleen, J., Sullivan, A., Yee, J., Wang, O., Gunawardane, P., Collins, I., and Davis, J. 2007. Synthetic Shutter Speed Imaging. *Computer Graphics Forum 26,* 3, 591–598.

[58] Trobin, W., Pock, T., Cremers, D., and Bischof, H. 2008a. Continuous Energy Minimization Via Repeated Binary Fusion . In *Proc. of European Conference on Computer Vision (ECCV'08)*. Vol. 5305. 677–690.

[59] Trobin, W., Pock, T., Cremers, D., and Bischof, H. 2008b. An unbiased second-order prior for high-accuracy motion estimation. In *Proc. of DAGM Symposium on Pattern Recognition*. Springer-Verlag, Berlin, Heidelberg, 396–405.

[60] Vaudrey, T., Rabe, C., Klette, R., and Milburn, J. 2008. Differences Between Stereo and Motion Behaviour on Synthetic and Real-World Stereo Sequences. In *Proc. of International Conference on Image and Vision Computing New Zealand (IVCNZ'08)*.

[61] Wang, H., Sun, M., and Yang, R. 2007. Space-Time Light Field Rendering. *IEEE Trans. Visualization and Computer Graphics 13,* 4, 697–710.

[62] Wedel, A., Cremers, D., Pock, T., and Bischof, H. 2009. Structure- and motion-adaptive regularization for high accuracy optic flow. In *IEEE International Conference on Computer Vision (ICCV'09)*. Kyoto, Japan.

[63] Wedel, A., Pock, T., Braun, J., Franke, U., and Cremers, D. 2008. Duality tv-l1 flow with fundamental matrix prior. In *Proc. of International Conference on Image and Vision Computing New Zealand (IVCNZ'08)*. 1–6.

[64] Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., and Debevec, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. on Graphics 24,* 3, 756–764.

[65] Werlberger, M., Pock, T., and Bischof, H. 2010. Motion Estimation with Non-Local Total Variation Regularization. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'10)*. IEEE Computer Society, 2464–2471.

[66] Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., and Bischof, H. 2009. Anisotropic Huber-L1 optical flow. In *Proc. of British Machine Vision Conference*. London, UK. to appear.

[67] Xiao, J., Cheng, H., Sawhney, H., Rao, C., Isnardi, M., and Corporation, S. 2006. Bilateral filtering-based optical flow estimation with occlusion detection. In *Proc. of European Conference on Computer Vision (ECCV'06)*. Vol. 1. Springer, 211–224.

[68] Xu, L., Chen, J., and Jia, J. 2008. A Segmentation Based Variational Model for Accurate Optical Flow Estimation. In *Proc. of European Conference on Computer Vision (ECCV'08)*. Springer, Berlin, 671–684.

[69] Zach, C., Pock, T., and Bischof, H. 2007. A duality based approach for realtime TV-L^1 optical flow. In *Proc. of DAGM Symposium on Pattern Recognition*. Vol. 4713. 214–223.

[70] Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgada, A., Rosenhahn, B., and Seidel, H.-P. 2009. Complementary optical flow. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, D.Cremers, Y.Boykov, A.Blake, and F.R.Schmidt, Eds. Lecture Notes in Computer Science, vol. 5681. 207–220.

[71] Zitnick, C., Jojic, N., and Kang, S. B. 2005. Consistent Segmentation for Optical Flow Estimation. In *IEEE International Conference on Computer Vision (ICCV'05)*. Vol. 2. 1308–1315.